論　文

２次元射影像からの３次元物体の認識と類別
- モジュール構造を用いた教師なし学習モデル -

正　員　鈴木　敏†　　正　員　安藤　広志†

3D Object Recognition from 2D Views
- An Unsupervised Learning Model Using Modular Structure -

Satoshi SUZUKI†, *Member*　and　Hiroshi ANDO†, *Member*

あらまし　本論文では２次元射影像のみから複数の３次元物体の認識・類別を行う神経回路モデルを提案する．本モデルは複数のモジュールからなり，モジュール間で競合を行いながらそれぞれのモジュールで入力射影像の圧縮・復元を学習する．その結果，各モジュールはそれぞれ一つの物体の射影像のみを復元できるようになり，その復元精度から物体の類別を行うことが可能となる．この過程では物体のラベル等の教師信号は不要である．さらに，本モデルでは入力特徴を限定していない．すなわち，入力情報を画像濃淡値としても，各特徴点の座標としても同様のネットワーク構造で扱うことができる．本論文では本モデルの詳細を説明するとともに，計算機実験の結果も合わせて提示する．この実験結果は，視点方向によらない３次元物体の認識が教師信号なしに可能であり，各モジュールの内部表現（圧縮表現）は視点方向に等価であることを示している．これは２次元情報のみからの３次元情報の推定を意味している．また，脳内の神経結合と比較して，情報の圧縮・復元過程は２つの視覚領野間の双方向結合に，モジュール間の競合は視覚領野内の水平方向の結合に対応づけて考えることができる．

キーワード　３次元物体認識，モジュール構造，教師なし学習，情報圧縮・復元，競合学習，双方向結合

1. まえがき

我々の脳は外界に存在する３次元物体をその２次元網膜像から認識・類別している．しかしながら，網膜上に投影される２次元射影像はその物体を見る方向に依存して大きく変化する．従って，我々の脳は視点が変化したときの異なる２次元情報からそれらが同一物体かどうかを判断できる機能を有していると考えられる．脳の機能を考える上で，この物体認識の機構と原理を明らかにすることは重要な課題である．

この問題に関して，心理学，生理学，計算理論など様々な分野で研究が進められている．例えば心理学では，３次元物体認識の視点依存性について検討がなされている．すなわち，物体認識は視点に依存した２次元射影像の記憶によるとする立場[5]と，視点に依存しない特徴により行われているとする立場[1]とに分かれて議論が行われている．また，生理実験の結果から認識のために使われていると考えられる脳の部位間の神経結合が明らかになりつつある．例えば，低次視覚野から高次視覚野にかけて双方向の結合の存在が確認されている．したがって，このような領野間でモジュール群が相互に関連しあうことで視覚処理が行われている可能性がある．しかしながら，このような双方向結合により何が計算されているのかについてはまだほとんどわかっていない．本論文では，計算理論の立場から，モジュール構造および双方向結合を利用した視覚認識に関するモデルを提案する．

これまでに物体認識に関する計算モデルはいくつか提案されている．例えば，Marrは画像から３次元物体の形状を復元し，物体中心の座標系により記述することで物体を認識するという考えを示したが[10]，これを具体的に実現する計算機構は提案されていない．また，実際に認識課題の遂行に成功したモデルはそのほとんどが教師あり学習を前提としており，２次元射影像以外に物体のラベルなどの教師信号を必要としているか[15]，あるいは，物体毎に学習を行う必要があった[18]．しかしながら，人間の視覚認識過程では網膜に写る２次元射影像のみを入力として３次元物体を類別（クラスタリング）できる可能性がある．すなわち，

† （株）ATR人間情報通信研究所，京都府
ATR Human Information Processing Research Laboratories, Kyoto, 619-02, Japan

教師なし学習による認識を考える必要がある．本論文で提案するモデルは情報の圧縮・復元およびモジュール間の競合を利用して教師なし学習を実現している．情報の恒等写像を学習するため，２次元射影像のみを入力とした３次元物体の認識・類別が可能である．

以下，２．では本論文で提案する３次元物体の認識・類別モデルについて詳細を述べる．３．では，折れ曲がった針金状の３次元物体を生成して複数の物体の類別実験を行なった．その方法と結果について説明する．ここでは入力として射影像の各頂点の座標を用いた実験と角度を用いた実験を行っている．さらに，４．では今回の実験課題について，従来の手法との比較・検討を行い，このモデルの有効性を示す．さらに，脳のモデルとして生理学的見地からの検討も行う．最後に５．で，本稿の結論と今後の課題を簡単にまとめる．

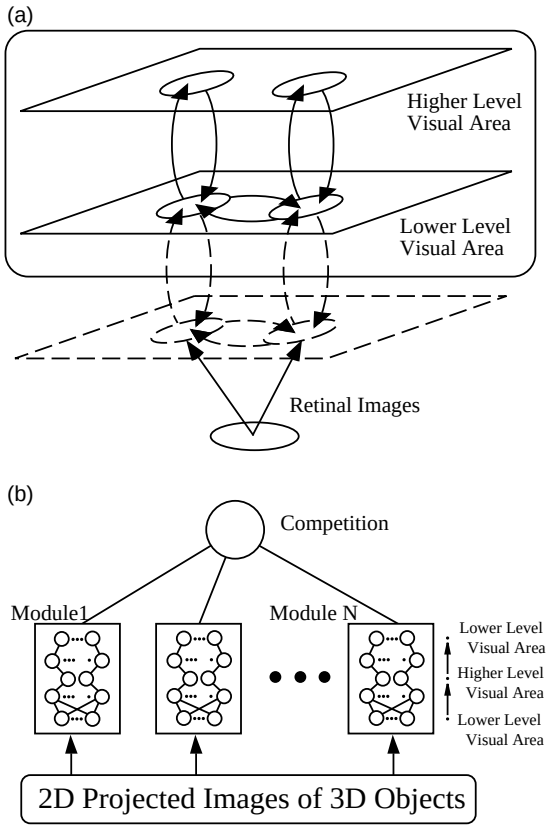

図1 (a) 脳内の認識過程の神経結合モデル，(b) ３次元物体認識・類別のためのネットワークモデル
Fig. 1 (a) A neural connection model for recognition in the brain, (b) The proposed network model for 3D object recognition.

2. 認識・類別モデル

本章では３次元物体を２次元射影像のみから認識するための教師なし学習モデルについて，脳内神経結合との比較を交えながら詳細に説明する．

2.1 モデルの概要

脳内の情報処理は双方向の神経結合によりつながれた様々な領野を通して階層的に行われていることが知られている．また，各領野内には水平方向の結合の存在が知られている．視覚認識に関係すると考えられている領野間でも同様に双方向結合が存在し，階層的な情報処理過程が考えられている．図１(a)はそのような神経結合の概略図であり，この中で一組の領野間の双方向結合を考える．このとき，低い領野→高い領野，高い領野→低い領野の結合をそれぞれ３層のネットワークで表わし，これらをまとめた双方向結合を５層の砂時計型ネットワークとして置き換えたモデルが図１(b)である．ここで，低い領野における水平方向の結合は，以下に説明するようにモジュール間の競合過程として表現されている．

この図１(b)が本論文で提案する３次元物体認識モデルである．このモデルは複数のモジュールとその出力を統合する部分から成り立っている．全てのモジュールは同一の構造を持っており，第２，第４層にシグモイド関数を持つ５層の砂時計型ネットワークである．また，第３層は次元が低く抑えられており，はじめの３層で入力情報の圧縮，次の３層で復元を行う．

各モジュールは入力情報の恒等写像を学習するため，このモデルでは物体のラベルなどの教師信号を必要としない．また，モジュール第３層の次元を低くしている（次元圧縮の拘束条件）ため，学習が進むにつれて一つのモジュールが一つの物体のみを復元できるようになることが期待される．

学習時は，任意の物体，任意の視点から生成される入力射影像が各モジュールに等しく与えられる．各モジュールはこの入力情報の圧縮・復元を学習するが，モジュール間の競合のため，入力を復元できるのは一つのモジュールのみとなる．さらに，物体の射影像の連続性から(注)，学習が進むにつれて一つの物体は一つのモジュールのみで復元が可能となる．

モジュールに復元可能な射影像が与えられると初め

注：本モデルで用いる入力特徴は３次元物体の姿勢表現と一意に対応するものとしている．一般に３次元物体の姿勢は連続的に変化するため，入力射影像を表わすベクトルもまた連続な値を示す．

の３層で非線形の情報圧縮 F を行い，次の３層で圧縮された情報の復元 F^{-1} が計算される． F^{-1} は F の逆変換を意味し，入力となる２次元射影像の集合を $\Omega \subseteq R^N$，圧縮表現の集合を $\Phi \subseteq R^M\ (N > M)$ とすれば，情報の変換は

$$\Lambda \xrightarrow{G} \Omega \xrightarrow{F} \Phi \xrightarrow{F^{-1}} \Omega$$

と書き表せる．ここで $\mathbf{\Lambda}$ は３次元物体の姿勢を表わす表現の集合， G は３次元物体の姿勢から射影像を生成する変換を表わしている．復元されたそれぞれの情報は入力と比較され，その復元精度 f によりモジュール間で競合を行う．最も復元精度の高いモジュールのみが結合を強化され，より精度の高い復元を行えるように学習が進められる．

i 番目のモジュールの復元精度 f_i はsoftmax関数により，

$$f_i = \exp\left[-\left\|y^* - y_i\right\|^2\right] \Big/ \sum_j \exp\left[-\left\|y^* - y_j\right\|^2\right] \quad (1)$$

で表わされる．ここで， y^* は入力， y_i は i 番目のモジュールの出力を示す．この式により，復元精度 f はモジュール間の違いを強調した形で相対的に表現されている．すなわち，合計が「１」になるように正規化されている．例えば，一つのモジュールの復元精度が「１」になれば他のモジュールの復元精度は全て「０」になる．また，中間層の次元を圧縮することで，個々のモジュールが複数の物体の復元を行うことを抑制している．よって，ある物体の射影像が入力として用いられたとき，学習が進むにつれて一つのモジュールの復元精度のみが大きな値を示すようになることが期待される（2.4参照）．

本モデルでは各モジュールに５層の砂時計型ネットワークを適用しているが，砂時計型ネットワークは３層の場合（変換が線形の場合）には主成分分析に近い計算が行われることが知られている[3][14]．しかし，本論文で示す実験条件の下では線形変換による圧縮・復元では十分な結果を期待できない．これは，一般に線形変換では圧縮情報からの復元が不可能な場合がほとんどであるためである．また，圧縮過程においても情報量を落とさない線形の圧縮が不可能な場合も多い．したがって，圧縮・復元のそれぞれを３層の非線形変換で行う５層のネットワークを用いる必要がある．

2.2 圧縮表現の次元とユニット数

各モジュールの第３層のユニット数，すなわち圧縮表現の次元について考える．

入力情報は，復元が可能な範囲内で圧縮の次元をできる限り低くすることが望まれる．これにより一つのモジュールで複数の物体を復元することをより困難にできる．復元可能な最低の次元は，多くの場合，３次元物体の姿勢を表現するパラメタの数（自由度の次元）に等しい．ただし，このパラメタの数と第３層のユニット数が一致しない場合が存在する．すなわち，周期性を持つパラメタのみで物体が表現される場合，ユニット数はパラメタ数より１つ多く必要となる（付録１参照）．

例えば，視点方向を固定して，線形の拡大・縮小をする静止した物体の射影像を入力とすれば第３層は１つのユニットで十分であるが，一つの軸に対して360°回転する剛体の射影像を入力として用いれば第３層のユニット数は２つ必要である．

2.3 入力特徴に関する普遍性

本モデルでは入力として用いる特徴を３次元物体の姿勢表現と一意に対応する限り自由に選ぶことができる（付録２参照）．

例えば，入力特徴は射影像の各頂点の座標であっても角度であっても良いし，濃淡画像ベクトルであってもよい．ただし，各画素が２値で表わされるビットマップ画像などは上記の条件を満たさないので入力特徴として用いることは難しい．

2.4 評価関数

本モデルでは，一つのモジュールが一つの物体の射影像のみを復元することが，期待される最適な状態である．これを学習により実現するために，ネットワークは評価関数 $\ln L$ を最大化することで最適化を行う． L は次の式で表わされる関数である．

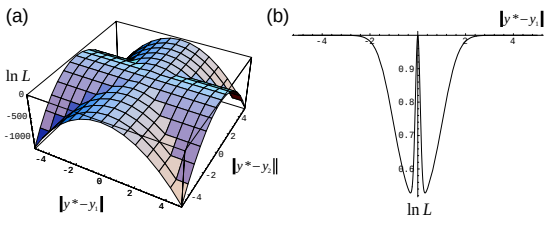

図2 モジュール２つの場合の復元誤差と評価関数の関係(a)と，片方の誤差を「０」としたときの断面(b)

Fig. 2 (a) Objective function for two modules, (b) cross section at $\left\|y^* - y_2\right\| = 0$.

$$L=\frac{\sum_i \exp\left[-\alpha\left\|y^*-y_i\right\|^2\right]}{\sum_j \exp\left[-\beta\left\|y^*-y_j\right\|^2\right]},\quad (\alpha>\beta>0). \qquad (2)$$

ここで，α,βはそれぞれ定数である．この評価関数はある入力に対し，一つのモジュールの復元誤差のみを「0」に近付け，他のモジュールの復元誤差を大きくするように作用する．すなわち，モジュールを競合させる機能を持つ．

図2はこの評価関数の性質を示すもので，例としてモジュールが2つある場合のモデルについて$\alpha=100,\beta=1$として示してある．(a)はそれぞれのモジュールにおける復元誤差$\|y^*-y_i\|$の組に対して評価関数の値$\ln L$を縦軸に示している（実際には復元誤差は正の値のみをとるが参考のため負の場合も示してある）．(b)は一つのモジュールの復元誤差を「0」で固定した場合の他方の復元誤差と$\ln L$との関係を示している．即ち，(a)において一つのモジュールの復元誤差が「0」の場合の断面を示している．これらの図から，この評価関数を最大化することによる競合の作用により，両方のモジュールの復元誤差が共に大きい場合は一方の復元誤差が小さくなる方向へ学習が進み(a)，片方の復元誤差が小さい場合には他方は，復元誤差がはじめから小さい場合を除いて，復元誤差が大きくなる方向へ学習が進むことがわかる(b)．従ってこの評価関数を最大化することにより，一つのモジュールは一つの物体のみを復元するように学習が進むことが期待される．

この評価関数を最大にする一つの解として，複数のモジュールが同時に復元誤差「0」を示す場合も考えられる．しかし，前述した次元圧縮の拘束条件により多くの場合一つのモジュールは一つの物体のみ復元可能となるので，同時に復元誤差「0」となるような解に到達することは難しくなっている．

3. 計算機実験

本モデルの有効性を示すため，計算機による類別の実験を行う．類別結果の詳細を示すとともに，各モジュールの中間層での内部表現および汎化の様子についても合わせて調べる．

3.1 入力情報の設定

実験では，入力物体として折れ曲がった針金状の3次元物体を用いた[15]．この物体は3次元空間

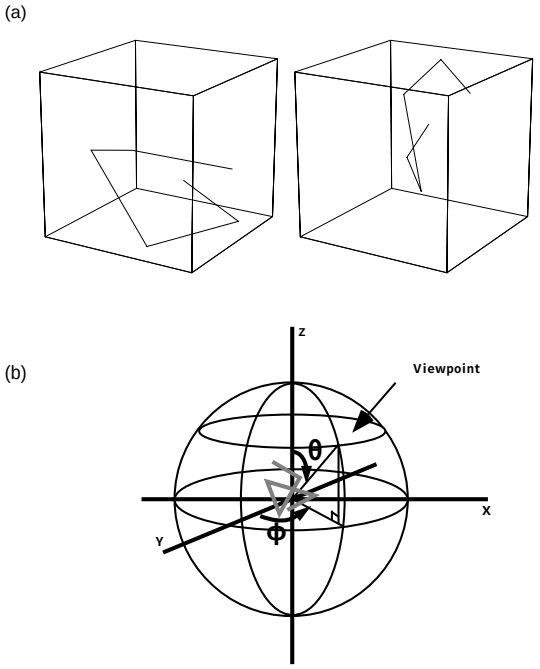

図3 (a)入力物体の例，(b)視点方向と射影像の生成
Fig. 3 (a) Examples of 3D objects, (b) view directions and generation of 2D images.

$\{-1\le x,y,z<1\}$の範囲内にある任意の6点を直線で結んで作成した．図3(a)にこの3次元物体の例を示す．このようにして作成した物体を3種類用意して入力として用いた．

ネットワークへの入力は，これらの物体をさまざまな視点方向から見たときの射影像とする．ここで，視点方向のベクトルは単位球面上で2つの方向角θ,ϕを用いて$(\sin\theta\cos\phi,\sin\theta\sin\phi,\cos\theta)$と表すことができる（図3(b)）．これを用いれば$\theta,\phi$の範囲を$\{0\le\theta\le\pi,0\le\phi<2\pi\}$とすることで全ての視点方向を表現できる（注）．

実際に入力特徴として用いたのは射影像の各点の座標ベクトルおよび各頂点角の余弦である．座標を入力特徴として用いる場合，6点からなる物体であるため，入力は12次元のベクトルとなる．この入力ベクトルは3次元座標ベクトルを要素とする行列をsとして

注：ここでは射影平面上での射影像の回転および位置ずれについては考慮していない．このため，視点方向を表現するパラメタは周期的パラメタおよび非周期的パラメタの2つである（付録1参照）．したがって，各モジュールの第3層のユニット数は2つで十分である．一般に，剛体物体が任意に回転運動をするとき，その姿勢はオイラー角の表現により，周期的パラメタ2つと非周期的パラメタ1つで表わされる．

$$
\begin{aligned}
t &= IR(\theta,\phi)\, s \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}
\begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}
\begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} s \\
&= \begin{pmatrix} \cos\theta\cos\phi & -\cos\theta\sin\phi & -\sin\theta \\ \sin\phi & \cos\phi & 0 \end{pmatrix} s
\end{aligned}
\qquad (3)
$$

により求められる．ここで I はXY平面への射影を表わす．この式は視点方向をZ軸方向にとった固定座標系で，物体をZ軸周りに回転させた後，Y軸周りに回転させたときの，XY平面への正射影を意味している．

一方，角度を入力として用いる場合，入力ベクトル u は４次元となり，i 番目の点の射影像上の位置ベクトルを t_i とすれば，

$$
u_i = \frac{(t_i - t_{i-1}) \times (t_i - t_{i+1})}{|t_i - t_{i-1}||t_i - t_{i+1}|} \bullet \mathbf{Z} \qquad (4)
$$

と，表わされる．ここで符号 $\times$ は外積，$\bullet$ は内積を表わす．また，$\mathbf{Z}$ はZ方向の単位ベクトルである．この表現によると，鏡像関係にある射影像は互いに符号が反対となる．

3.2 シミュレーション

入力物体を３種類としたため，３つのモジュールから成るネットワークの構成を用いる．各モジュールのユニット数は第２，４層を２０，第３層を２，第１，５層を入力ベクトルの次元と同数に定めている（2.2および3.1参照）．

学習は任意の射影像をネットワークに順次与え，圧縮・復元を繰り返すことで行う．このとき，物体のラベルなどの教師信号は与えない．

実験では射影像を完全に任意に生成する方法と，視点範囲を連続的に拡大しながら射影像を生成する方法の２種類を行った．その結果，どちらも学習は収束したが，視点範囲を拡大する実験のほうが収束の速さが約１０倍であった（座標入力の場合）．よって本論文では視点範囲を拡大する場合の実験について詳細を述べる．

視点範囲を拡大する実験では，任意の物体の定められた視点範囲内において任意の視点方向への射影像を生成し，ネットワークに与えて圧縮・復元を学習させることを繰り返す．このとき，視点範囲は連続的に拡大するように変化させる．すなわち，視点方向 θ,ϕ の範囲をそれぞれ $\pi/4, \pi/2$ の幅から徐々に拡大し，最終的に全視点方向を覆うように変化させる．また，３種類の物体の出現率は共に等しく，物体による学習の片寄りはない．学習は，評価関数の値 $\ln L$ を最大化することで行われる．今回の実験では，評価関数は式（２）で $\alpha = 100, \beta = 1$ とした．

また，学習していない視点方向に対してどの程度の汎化が起こるかを調べるため，限定した視点範囲内のみで視点方向を学習する実験も行い，学習範囲及びその周囲の類別結果などの様子を調べた．学習範囲は θ,ϕ をそれぞれ $\pi/4, \pi/2$ の幅でとった．

今回の実験では最適化のアルゴリズムは最急降下（上昇）法（誤差逆伝搬法）を用いたが，収束を早めるために共役勾配法などを用いることも可能である．

3.3 結果

以下，計算機による実験の結果を示す．テストは学習終了後，ネットワークの重みを固定して行っている．

はじめに，図４に学習曲線を示す．縦軸には物体類別の正答率，横軸には学習時間（繰り返し回数：学習回数）が示してある．ここで，一回の学習とは一つの物体の一つの射影像が入力としてネットワークに与えられたときの一回の重みの更新とする．テストは全視点範囲から等間隔に選ばれた2500の視点方向で行われている．この図では座標入力の場合と角度入力の場合が示されている．座標入力では学習回数が10^3で７割，10^4で８割の正答率を示している．100%の正答率を得るには10^6程度の学習回数が必要となり，多少の時間が必要となる．また，座標入力の場合ほど精度は良くないが，角度入力でも学習回数が増えるにしたがって復元精度がよくなっている．このように，異なる入力特徴を用いた場合でも同じ構造のネットワークで同様の

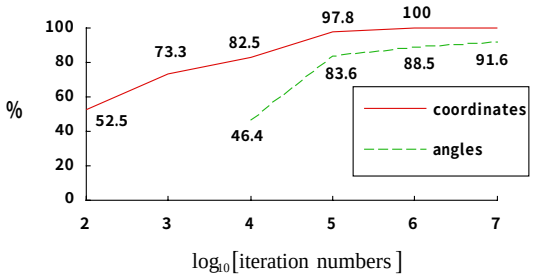

図4 学習曲線
Fig. 4 Percent correct of classification.

類別結果を得ることができる．

以下，座標入力の場合の結果についてさらに詳細を示す．

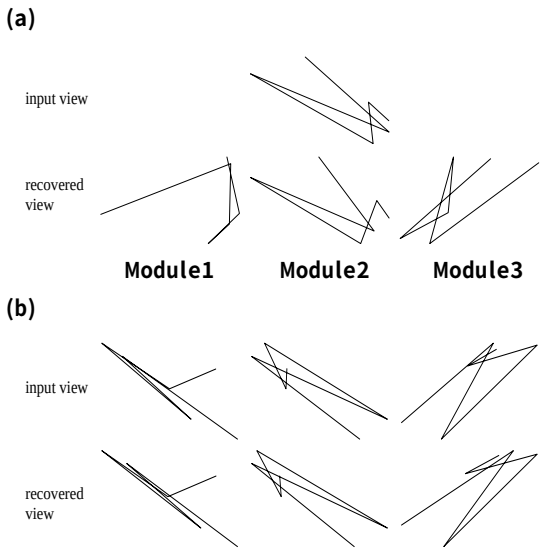

図5 入力／復元像，(a) ある射影像に対する各モジュールの復元像，(b) ある物体の射影像と対応するモジュールでの復元像
Fig. 5 Input/recovered views: (a) a projected view and its recovered views by three modules, (b) projected views of an object and their recovered views by the corresponding module.

まず，100%の正答率が得られたときのネットワークについて詳細に調べる．図５(a)はある物体(Object3)のある射影像を入力として用いたときの各モジュールでの復元の様子を示している．この物体に関しては２番目のモジュール(Module2)が最も精度の高い復元を示しており，Module2がObject3に対応していることがわかる．(b)はObject3の他の視点からの射影像を入力としたときの，Module2での復元像を示している．上段の入力像と下段の復元像を比較すると，両者は非常に似た形をしており，モジュールはかなり正確に射影像を復元していることがわかる．

このことを全視点範囲にわたって示しているのが図６である．(a)はObject3を入力として用いたときのモジュールごとの相対的な復元精度を示している．この図は全視点方向（θ, ϕ）に対する復元精度 f（式(1)）の値を表示しており，各視点方向からの射影像一つ一つに対する類別結果を意味している．この図から全視点方向で一つのモジュール（Module2）の復元精度だけが「１」に近い値を示していることがわかる．これは，一つの物体(Object3)を入力として用いたとき，その視点方向に関わらずそれらが同じ物体であると判断したことを意味しており，この物体に対する類別は視

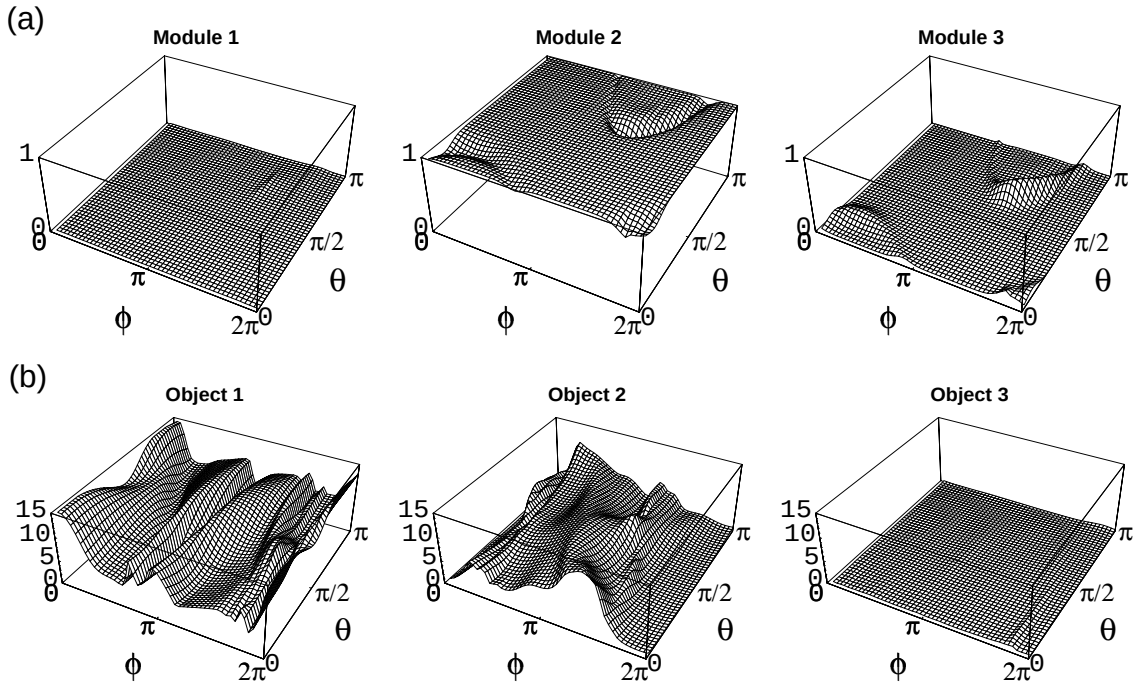

図6 (a) 復元精度のモジュール間での比較，(b) 物体による復元誤差の違い
Fig. 6 (a) Relative recovery accuracy for each module, (b) recovery errors for each object.

点方向によらず，完全に行なわれているといえる．他の物体を入力として用いたときも同様に一つのモジュールの復元精度のみが「１」に近い値を示す．

また，この第２番目のモジュールにそれぞれの物体の射影像を入力した場合の入力像と復元像との誤差 $\|y^*-y_i\|^2$（各頂点座標の２乗誤差の和）の大きさを表したのが図６(b)である．このモジュールで復元可能な物体（Object 3）を入力として用いた場合の復元誤差の大きさと比較して，他の物体を入力とした場合の復元誤差は相対的に大きいことがわかる．これは次元圧縮の拘束条件のために，一つのモジュールで復元できる物体が一つだけに限られていることを示している．すなわち，各モジュールにはそれぞれに対応する物体が存在し，それぞれの射影像のみを復元できる．

このように図６(a)(b)から，入力物体とモジュールは１対１の関係になっており，類別は完全に行われていることがわかる．

さらに，圧縮された表現（モジュールの第３層の出力）と視点方向 θ,ϕ との関係を調べた．結果を図７に示す．(a)はある物体の射影像の圧縮変換による像を表わしている．周期的パラメタ ϕ と非周期的パラメタ θ の２つのパラメタからなる３次元物体の像であるため穴のあいた円盤状の表現になっている．このことから，視点方向 (θ,ϕ) と圧縮表現 $(\mathrm{unit1},\mathrm{unit2})$ の対応関係がほぼ１対１であることがわかる．(b)は圧縮過程による視点方向 (θ,ϕ) に対する第３層の各ユニット unit1, unit2 の出力であり，(c)はその逆，すなわち，第３層の出力の組 $(\mathrm{unit1},\mathrm{unit2})$ に対して視点方向の各要素 θ,ϕ を表示したものである．(c)の図から第３層の出力に対して θ 及び ϕ が一意に対応しており，圧縮表現が求まれば視点方向はほぼ一意に決まることがわかる．逆に(b)は，視点方向に対して第３層の出力が圧縮過程（第１層から第３層への一価関数）により一意に決まることを示している．よって(b)(c)からも視点方向と圧縮表現との間には１対１の対応関係が成り立っていることがわかる．ちなみに，(a)は(c)の各図を真上から見た図である．以上の結果は，各モジュールの第３層ではそれぞれに対応する物体に関して，視点方向に等価な表現が得られていることを示している．言い替えれば，各モジュールにおいて物体固有の圧縮・復元過程が学習により獲得されたともいえる．この結果はモデルにおいて，第３層のユニット数を３次元物体の自由度に合わせていることから予想された通りの結果である（付録１参照）．

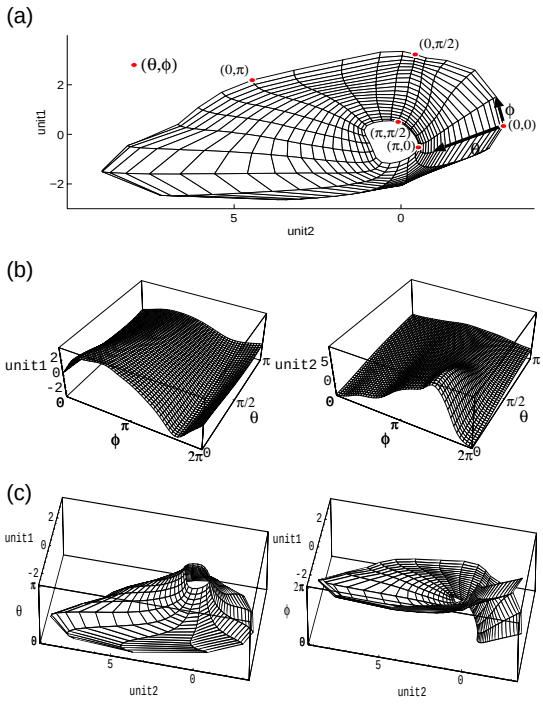

図7 視点方向と圧縮表現の関係, (a) 物体射影像の圧縮表現，(b)全視点方向に対する第３層の出力（圧縮表現の各要素）, (c)圧縮表現に対する視点方向の各要素

Fig. 7 Relation between view direction and compressed representation: (a) compressed representation, (b) output in the 3rd layer as a function of view direction, (c) view direction as a function of output in the 3rd layer.

図８には汎化の様子が示されている．中心の四角く囲まれている部分が学習範囲であり，テストは全視点領域で行われている．(a)は一つの物体を入力として用いたときの，その物体に対応するモジュールでの復元精度を表している．学習範囲を中心として広い範囲に渡って正しい類別が行われており，さらに遠くはなれるにしたがって徐々に復元精度が下がっている．一方，(b)ではこのときの入力像と復元像との誤差が示されており，学習範囲からはなれるに従って復元誤差は大きくなっている．これらの結果は，学習領域の外でも近い範囲では物体の認識が可能であることを示すものである．

また，汎化の課題については，視点方向を離散的にとった限られたデータセットによる補間の学習実験も行った．その結果，各視点方向を４等分にした合計１６点からなる視点方向の組を用いた実験で，10^5回の学習回数で約９０％の正答率を得ている（図４と比較参

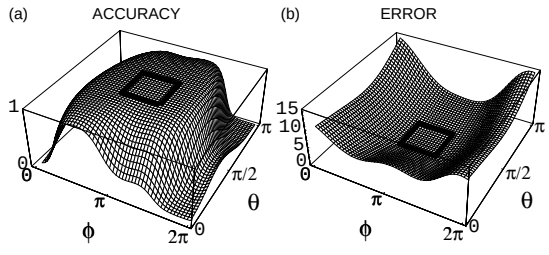

図8 汎化, (a)復元精度, (b)復元誤差
Fig. 8 Generalization: (a) recovery accuracy, (b) recovery error.

照）.

4. 考察

4.1 工学的モデルとしての検討

今回の実験で用いた類別課題を従来手法と比較するため，従来の代表的なクラスタリング手法を用いて同様の実験を試みた.

線形分離の可能性 今回の実験で用いた類別課題を単純パーセプトロンによる教師あり学習で試みた．その結果，識別誤差を「0」に近づけることはできなかった．この結果は今回行った類別課題は線形分離可能ではないことを示している．なぜなら，パーセプトロンの収束定理によれば，「識別課題が線形分離可能であるならば，有限回の学習でかならず正しい解に収束する」からである[2][13]（注）．したがって，この課題を線形の手法のみで解決することは不可能である.

k-means法との比較 従来の代表的なクラスタリング手法であるk-means法[4]を用いて同様の実験を行った．k-means法は入力空間での距離によるボロノイ分割を用いた類別手法である．その類別結果を図9に示す．これは計算機実験におけるテストと同様に，各物体2500枚の射影像を用いた類別実験の結果である．この図からクラスと物体の1対1の対応関係は見られず，一つのクラス内に複数の物体が混ざっていることがわかる．この結果は従来の入力空間の距離を直接用いた分割手法では今回行った類別課題を遂行することが困難であることを示している.

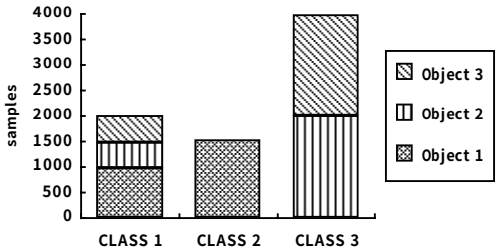

図9 k-means法による類別結果
Fig. 9 Clustering results using k-means method.

類似した手法との比較 村瀬とナイヤーは主成分分析を用いて固有空間上で多様体を構成するモデルを提案し，物体を視点方向に依存せず認識することに成功しているが[11][12]，このモデルはわれわれの提案するモデルにおいて各モジュールを3層（線形の圧縮・復元）にし，中間層の次元を多めに設定した場合のモデルに類似している（2.1参照）．しかし，村瀬らの提案するモデルでは情報圧縮時の距離を測っているのに対し，われわれのモデルでは復元した情報と入力との距離を測っているという点で違いがある．そのため村瀬らのモデルでは，圧縮次元における多様体の構成・補間・距離計算に煩雑な手続きを必要としている．さらに村瀬らのモデルでは，物体の類別は教師あり学習に基づいており，物体ごとにラベルを必要としている.

Jacobsらによるモジュール学習法との比較 Jacobsらはモジュール間の競合を用いて入力空間を直接分割するモジュール学習モデルを提案している[6][7]．このモデルに5層の砂時計型ネットワークを適用することで，今回の実験と同様の課題を行うことができる[16][17]．これらの結果からこのような類別課題に関しては我々の提案するモデルがわずかに優位であることがわかる．この違いは，彼等のモデルでは入力空間を直接分割する非線形変換の獲得のためにモジュールが利用されるのに対し，われわれのモデルでは各モジュールが物体の認識に直接利用される点に起因すると考えられる.

4.2 脳のモデルとしての検討

川人と乾は視覚大脳皮質の構造と機能に基づき，その計算機能を推測している[8][9]．その中で初期視覚野と高次視覚野の間の双方向結合を光学と逆光学として位置づけており，逆光学は2次元画像から3次元空間での情報の記述を取り出す手段であると推測している.

一方，われわれの提案するモデルでも圧縮・復元過程を異なる領野間の双方向結合として想定することが

（注）パーセプトロンの収束定理は2分割問題に関するものであるが，3分割の学習も2分割学習の組み合わせにより可能となる．すなわち，出力層のユニットを2つにすればよい．このとき出力層の各ユニットは異なる2分割問題を独立に学習するため，収束定理は同様に成り立つことになる．多分割問題でも同様である.

でき，圧縮表現として取り出される情報は視点方向という３次元空間での情報である．従って，本モデルは川人らのモデルを拡張し，認識・類別機能を実現したものとみなすこともできる．

また，脳の視覚情報処理では初期，中期の過程で様々な特徴が取り出され，高次機能へ入力として送り出されていると考えられる．この点においても，本モデルは特定の入力特徴を前提としていないため，初期過程で取り出される様々な特徴に対して柔軟に対応できる利点がある（付録２参照）．

さらに，視点方向に等価な内部表現は物体の姿勢推定などに利用でき，運動制御系への入力としても役に立つものと考えられる．

５．むすび

本研究では，モジュール構造を用いたネットワークモデルを提案し，３次元物体の認識･類別課題に対してその有効性をシミュレーションで検討した．その結果，次のことが明らかになった．第一に，複数の３次元物体の認識･類別が教師なし学習で実現できた．このときに，入力特徴によらず同一のネットワーク構造で物体の類別が可能である．第二に，各モジュールの中間層に圧縮表現として，視点方向に等価な表現が形成された．この結果は２次元情報から３次元情報の推定を示すものである．

しかし，今回の実験では学習のための時間が十分に必要であり，モジュール数を物体数に合わせて実験を行っている．これらの点が問題点として指摘できるが，物体数が不明の場合でもモジュール数を十分多く用意しておけば同様に類別が行われると考えられる．また，学習のためにかかる時間は他の収束を早める学習アルゴリズムの適用や並列処理の導入による改善が期待できる．

今後の課題としては上記の改善点に加えて，さらに様々な入力特徴を用いた実験を行い本モデルの有効性を確認したい．

謝辞 ATR人間情報通信研究所の川人光男室長には貴重な助言を頂き，深く感謝します．また，有意義な討論をして頂いたMITのT.Poggio教授に感謝します．

文献

[1] I. Biederman, "Human image understanding: recent research and a theory", computer vision, graphics, and image processing, vol.32, pp.29-73, 1985.

[2] H.D.Block, "The perceptron, a model for brain functioning I". Rev. of Modern Physics, 34, 123-135, 1962.

[3] D. DeMers and G. Cottrell, "Non-linear dimensionality reduction", Advances in Neural Information Processing Systems 5, San Mateo, CA., pp.580-587, 1993.

[4] R. O. Duda and P. E. Hart, "Pattern Classification and Scene Analysis",John Wiley & Sons, NY., 1973.

[5] S. Edelman and H. H. Bülthoff, "Orientation dependence in the recognition of familiar and novel views of three-dimensional objects", Vision Research, vol.32, no.12, pp.2385-2400, 1992.

[6] R. A. Jacobs, M. I. Jordan, S. J. Nowlan and G. E. Hinton, "Adaptive mixtures of local experts", Neural Computation, vol.3, pp.79-87, 1991.

[7] R. A. Jacobs and M. I. Jordan, "Learning piecewise control strategies in a modular neural network architecture", IEEE Transactions on Systems, Man, and Cybernetics, vol.23, no.2, pp.337-345, 1993.

[8] 川人　光男, 乾　敏郎, "視覚大脳皮質の計算理論", 電子情報通信学会論文誌D-II, vol.J73-D-II, no.8, pp.1111-1121, 1990.

[9] M. Kawato, H. Hayakawa and T. Inui, "A forward-inverse optics model of reciprocal connections between visual cortical areas", Network, vol.4, pp.415-422, 1993.

[10] D. Marr, "Vision", 乾　敏郎　安藤　広志 訳, ビジョン－視覚の計算理論と脳内表現－, 産業図書, 1982.

[11] H. Murase and S.K. Nayar, "Visual learning and recognition of 3-D objects from appearance", International Journal of Computer Vision, vol.14, pp.5-24, 1995.

[12] 村瀬　洋, シュリー　ナイヤー, "２次元照合による３次元物体認識－パラメトリック固有空間法－", 電子情報通信学会論文誌D-II, vol.J77-D-II, no.11, pp.2179-2187, 1994.

[13] N.Nilsson, "Learning Machines", 渡辺訳, 学習機械, コロナ社, 1967.

[14] E. Oja, "Data compression, feature extraction, and autoassociation in feedforward neural networks", Artificial Neural Networks, ed. T. Kohonen, K. Mäkisara, O. Simula and J. Kangas, pp.737-745, Elsevier Science Publishers, B.V. North-Holland, 1991.

[15] T. Poggio and S. Edelman, "A network that learns to recognize three-dimensional objects", Nature, vol.343, pp.263, 1990.

[16] 鈴木　敏, 安藤　広志, "モジュール学習による３次元物体の認識と類別", 信学技法, NC93-62, pp.59-66, Dec. 1993.

[17] S. Suzuki and H. Ando, "Unsupervised classification of 3D objects from 2D views", Advances in Neural Information Processing Systems 7, MIT Press,1995.

[18] D. Weinshall, S. Edelman and H. H. Bülthoff, "A self-organizing multiple-view representation of 3D objects", Advances in Neural Information Processing Systems 2, San Mateo, CA, pp.274-281, 1990.

[19] 横田　一郎, "群と表現", 基礎数学選書, 裳華房, 1973.

付録

1.　圧縮時の次元

圧縮時の次元，すなわち各モジュール第３層のユニット数について考察する．ここでは視点を固定して物体を変化（回転，拡大・縮小など）させるものとして議論を進める．

入力として用いる1つの3次元物体に関して，（環境の表現，例えば光源方向なども含めた）姿勢表現をユークリッド空間上に表わし，その集合を$\Lambda \subseteq R^L$，2次元射影像の集合を$\Omega \subseteq R^N$，圧縮表現の集合を$\Phi \subseteq R^M$とする．3次元物体の姿勢表現から2次元射影像への写像G，情報圧縮の写像F，復元の写像F^{-1}の関係は

$$\mathbf{\Lambda} \xrightarrow{G} \mathbf{\Omega} \xrightarrow{F} \mathbf{\Phi} \xrightarrow{F^{-1}} \mathbf{\Omega}$$

と，表わされる．

3次元物体が回転などを含む場合，同一の姿勢を表現するのにユークリッド空間上では同じ座標系を用いても複数の表現が可能である．言い替えれば，3次元物体のある姿勢からユークリッド空間上の姿勢表現への写像は1対多である．すなわち，$\mathbf{\Lambda}$の元$x \in \mathbf{\Lambda}$をL個のパラメタで$x=(x_1,\cdots,x_i,\cdots,x_L)$と表わすとき，あるパラメタ$x_i$に関して，ある値$\alpha \neq 0$が存在し，

$$G(x_1,\cdots,x_i,\cdots,x_L)=G(x_1,\cdots,x_i+n\alpha,\cdots,x_L)\in\mathbf{\Omega},\quad n\in Z$$

を満たす場合がある．例えば，このようなx_iを周期的パラメタと呼ぶことにすると周期的パラメタには物体の回転の他に光源方向の変化なども含まれる．また，非周期的パラメタは平行移動，明暗強度，範囲の限られた回転など，有限の値で表わされるものが挙げられる．

このような周期性を持つパラメタは1次元球面（円周）S^1上に連続に射影でき，S^1を複素平面上にとることで姿勢表現を複素空間上に一意に表わすことができる．例えば，$m(\leq L)$個の周期的パラメタを持つ物体を考えた場合，それぞれの周期的パラメタを複素平面上のS^1へ写す写像による$\mathbf{\Lambda}$の像を$\mathbf{\Lambda}'$とすると，$\mathbf{\Lambda}'$は$m(\leq L)$個のS^1と非周期的パラメタのみからなる部分空間内の集合Γとの直積集合$S^1\times S^1\times\cdots\times\mathbf{\Gamma}$として表わせる．すなわち，全ての姿勢表現は$\mathbf{\Lambda}'$で表わせる．

いま，復元可能の条件の下で最小となる圧縮次元Mを考える．復元可能であるためには写像$F:\mathbf{\Omega}\to\mathbf{\Phi}$が全単射(1:1,上への写像)であることが必要十分である．ここで，写像$G':\mathbf{\Lambda}'\to\mathbf{\Omega}$が全単射であると仮定する．すなわち，射影像から物体の姿勢が一意に定まる場合について考える．このとき写像$F\circ G':\mathbf{\Lambda}'\to\mathbf{\Phi}$もまた全単射となる．

Lを自由度の次元と等しくとると，Lが物体の姿勢を表現できる最小の次元であることから，圧縮時の次元$M\geq L$がFが全単射であるための必要条件となる．特に，姿勢表現が周期的パラメタのみからなる場合には，$\mathbf{\Lambda}'$はL個の1次元球面の直積集合$S^1\times S^1\times\cdots\times S^1$となり，その像は$L$次元トーラスを構成する[19]．このため，$L+1$次元以上のユークリッド空間が必要であり$M\geq L+1$が必要条件となる．逆に，上記の条件を満たす場合には$\mathbf{\Phi}$から$\mathbf{\Lambda}'$への写像$G'^{-1}\circ F^{-1}$が常に一意であることを考えると，姿勢表現が非周期的パラメタを含む場合には$M=L$，周期的パラメタのみの場合には$M=L+1$とすれば十分である．

したがって，写像$G':\mathbf{\Lambda}'\to\mathbf{\Omega}$が全単射であれば中間層のユニット数を$L$（非周期的パラメタ含む場合）あるいは$L+1$（周期的パラメタのみの場合）とすることで復元可能な最小の次元に圧縮できる．

2. 入力特徴に関する普遍性

上記議論における圧縮時の次元（第3層のユニット数）の決め方は，入力特徴として何を用いるかには依存していない．すなわち，写像G'がどのような特徴を抽出するかには関係なく，全単射であることのみが復元可能のための条件となる．従って，われわれの提案するモデルでは異なる入力特徴を同様のネットワーク構造で統一的に扱うことができる．例えば，各頂点の座標や濃淡画像を入力特徴とすれば写像G'はこの条件を満たすが，逆写像G'^{-1}が一意ではないビットマップ（2値）画像などでは条件を満たさない．ただし，実際の圧縮・復元の精度は第2，4層のユニット数に依存する．

鈴木　敏

平2東大・教養・基礎科学卒．同年NTT入社．平4からATR人間情報通信研究所へ出向．3次元物体認識に関する計算モデルの研究に従事．脳研究への計算論的アプローチに興味を持つ．平6日本神経回路学会研究賞受賞．日本神経回路学会会員．

安藤　広志

1983年京都大学理学部卒（物理学専攻）．1987年（京都大学文学部修士過程修了・心理学専攻）．1992年米国MIT脳・認知科学科博士過程修了，Ph.D.．1992年ATR人間情報通信研究所研究員．1994年より同研究所主任研究員．視覚情報処理，心理物理学，計算論的神経科学の研究に従事．